# A STUDY OF INFRASPECIFIC GROUPS OF THE BALTIC EAST COAST AUTUMN HERRING BY TWO NEW METHODS BASED ON CLUSTER ANALYSIS

E. Ojaveer
Estonian Laboratory of Marine Ichthyology

Appendix 1
J. Mullat
Tallinn Technical University

Appendix 2
L. Võhandu
Tallinn Technical University

In the Baltic Sea the autumn spawning herring forms a smaller number of groups than the spring herring does. This is probably connected with the different location of their spawning grounds. Spawning grounds of the spring herring are concentrated in favorable sites near the coast (in gulf, estuaries, etc.) while between such spawning centers gaps occur usually. Contrary to it, in most parts of the Baltic spawning places of the autumn herring form a continuous chain situated in the open sea. Therefore, differences in environment conditions between the autumn spawning grounds of neighboring areas are small and in large districts the characters of the autumn herring do not reveal essential differences. For instance, there is no significant difference between the autumn herrings caught on various grounds off the Polish coasts (2,5). The autumn herring of the Swedish Baltic coasts can be divided into four groups (that of the Gulf of Bothnia, that of the Bothnia Sea, the herring of the Swedish east coast and that of the Swedish south coast), between which a gradual transition occurs [1].

Environmental conditions in various parts of the northeastern Baltic differ considerably. Spawning grounds of the autumn herring in these parts are disconnected by large areas where no spawning occurs. Therefore, differences between the herring groups

connected with various spawning grounds are relatively big here. Already Suvorov [8] discerned (mainly on the ground of the diameter) there several autumn herring groups: those of the Gulf of Riga, Ventspils and Liepaja, the autumn herring spawning on Neckman-ground (north of Hiiumaa) etc. Presently, on the ground of differences in stock fluctuations four autumn herring groups inhabiting the different parts of the north-eastern Baltic (Gulf of Finland, Gulf of Riga, Venspils-Saarema, Hiiumaa) are treated separately. However, taking into account that areas of these parts are rather limited and the fact that areas of theses parts are probably widely mixed, the actual number of stocks and their boundaries remained uncertain.

## Material and methods

For the identification of the stocks in the autumn spawning herring off the east coast of the Baltic, morphological characters of fish caught from Klaipeda up to the eastern part of the Gulf of Finland were examined. For comparison a sample was taken off the Swedish coast. A total of 441 adolescent fish having finished their second growth period were investigated. All the fish belonged to the moderate 1967 year-class. Therefore, the material was not influenced by possible differences in age composition between various samples and in different rate of sexual maturation in various year-classes. The samples were taken by trawl in December, 1969 and in April, 1970 at 12 stations off the Baltic east coast and off the south coast of Sweden (Fig. 1, Table 1). The material was fixed in formalin for 6-12 hours (depending on the size of the fish) and conserved in alcohol. To avoid selection, at each station a random sample of adolescent harrings caught by trawl was fixed. Differentiations between the autumn and spring herrings as well as the age determination was made by means of otoliths after the measurements of the plastic features had been taken. The following plastic characters were determined: $L$, $l_S$ (both within 1 mm), head length, head height, eye diameter, iterorbital space, antedorsal space,

Fig. 1. The map showing the autumn herring sampling places

anteventral space, anteanal space, maximum body height, minimum body height, length of the pectorial fin (all to within 0.1 mm). From meristic characters the number of vertebrae, pyloric caecae, gill-rakers and pectorial fin rays were counted. Besides the

Table 1

Numbers of autumn herrings caught in various sampling places
(see Fig. 1) by subgroups (according to Figure 2)

| No. of station | Subgroups | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1a | 1a′ | 1b | 2′ | 2″ | 2‴ | 2⁗ | 3b | 3a′ | 3a | |
| 1 | 62 | 26 | | | | | | | | | 88 |
| 2 | 15 | 1 | 11 | 9 | | 3 | | 1 | | | 40 |
| 3 | 5 | | 4 | | | | | | | | 9 |
| 4 | | | 6 | 10 | 2 | 3 | 11 | 3 | | | 35 |
| 5 | | | 2 | 5 | 1 | 1 | 1 | 4 | 1 | 1 | 16 |
| 6 | | | 4 | 11 | 8 | 13 | 2 | 8 | | 4 | 50 |
| 7 | | | | 7 | 5 | 6 | 1 | 10 | 3 | 7 | 39 |
| 8 | | | | 1 | 1 | | | 7 | 22 | 24 | 55 |
| 9 | | | 4 | 15 | 7 | 8 | 5 | 14 | 2 | 11 | 66 |
| 10 | | | | | | | | 1 | 2 | 7 | 10 |
| 11 | | | | 1 | 6 | 1 | | 8 | 1 | 3 | 20 |
| 12 | | | | | | 1 | | 6 | 1 | 5 | 13 |
| Total | 82 | 27 | 31 | 59 | 30 | 36 | 20 | 62 | 32 | 62 | 441 |

plastic meristic characters commonly used for similar purposes, otolith measurements were included, since several authors differentiate between herring groups by means of otoloth characters (6, 4, 3). The total otolith length, the length and width of the first growth zone and the first and second winter zone (on postrostum) were measured (to within 0.025 mm). Sex, total weight and weight less gut (to within 0.1 g) were determined.

Morphological, physiological and other characters of a population are interrelated and form complexes. It can be presumed that the analysis of the features chosen allows to differentiate theses complexes with accuracy corresponding to our present knowledge in this field. For this purpose an electronic computer was used. An original program was applied for the identification of clusters and examination of interdependence between them. The method is based on the graph theory (see appendix 1).

The results obtained by this method were checked by another method that is based on the Hadamard transforms (appendix 2). For this purpose the characters (attributes) of all specimen were divided into two parts: 1) otolith measurements; 2) length, weigh, head length, eye diameter, numbers of vertebrae, gill-rakers, pyloric caecae, pectoral fin rays; 3) the rest of plastic measurements. On the ground of every set of characters, plane coordinate of points, representing individual fish, were found by the computer.

## Results

The position of points (representing specimens) in multidimensional space, found by the computer for each herring on the basis of its characters, was transformed into a plane coordinate system provided there was no significant changes in distances between the points. On the scheme drawn on the basis of the location of the points on the plane (Fig.2) the existence of a number of groups can be ascertained. The condition of formation of groups was that the distance (the notion of distance is given in appendices) between the neighboring points within the limits of a group was considerably smaller

than the distance between neighboring groups. The groups are arranged in the form of two spirals both finishing in the origin of coordinates. These two spirals represent two large sets of herring groups, differences between which are the most significant. In the middle part of the right-hand spiral a relatively sparse subgroups (1a, Fig. 2) of this set of groups is situated. Both this group and the subgroup 1a′ closely connected to it, are comparatively homogeneous consisting mainly of the Swidish south coast herring (Table 1). In the lower (nearer to the origin of coordinates) part of the subgroup the number of points, representing herrings caught off Klipeda and Liepaja, increases. In the part of the spiral situated closer to the origin of coordinates a relatively dense subgroup 1b is situated. It consists mainly of herrings of the Southern Baltic.
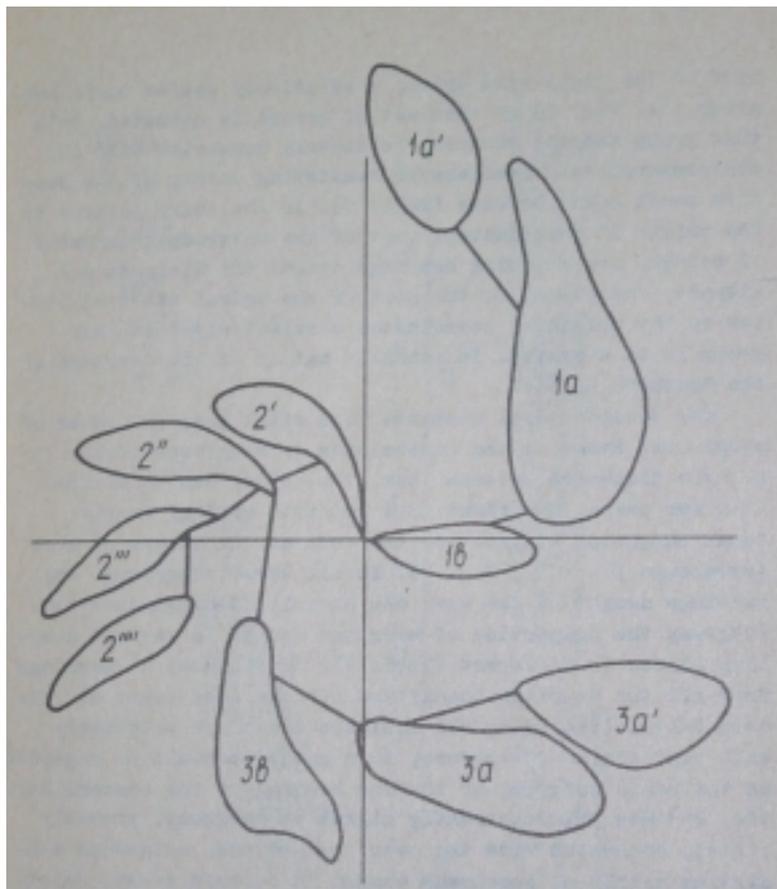


Fig. 2.   Positions of the autumn herring subgroups differentiated
by the method described  in Appendix 1

The second spiral consists of a still greater number of subgroups. Based on the composition of subgroups (Table 1) and the distances between them, the spiral can be divided into two parts. The first part consist of four nearly equal subgroups situated to the left of the origin of axes (subgroups 2′ - 2′′′′, Fig. 2). In all these subgroups the herrings caught in the open sea prevail although in every subgroup the proportion of herrings caught in various sampling places is different (Table 1). In subgroup 2′ herrings from all the sampling locations off the east coast of the open Baltic (including Klaipeda area) are relatively well represented. Therefore, this subgroup could be regarded as the basic subgroup of the sea herring of the Eastern Baltic. Besides this apparently migratory subgroup, probably closely connected with the open sea grounds, subgroup consisting mainly of specimen caught in certain areas, exists. In subgroup 2′′′ the herrings sampled in the vicinity of the Irben and Soela Sounds are dominating while group 2′′′′ contains chiefly specimen caught of Ventspils. Subgroup 2′′ seems to be a transitory group, consisting mainly of herrings caught in the transition areas between the sea herring and the gulf herring (the areas off the Irben and Seola Sounds, the middle part of the Gulf of Finland).

The second part of the spiral consists of two subgroups (3a and 3b, Fig.2) relatively well separated from each other. The dense basic subgroup (3a) situated relatively far from the origin axes, consists mainly of the herring caught in the gulf of Riga, in the Gulf of Finland and in the vicinity of the Soela Sound (Table 1). Subgroup 3a′ is comparatively close and similar to this basic subgroup (the herring of the Gulf of Riga prevails here). Subgroup 3b is located at a greater distance from the basic subgroup. Its composition shows that this is a transient unit (it consists mainly of specimen caught on the transition grounds between the areas inhabited by the sea herring and gulf herring). As regards the composition subgroup 3b is similar to subgroup 2′′. Morover, there id a direct link between them (Fig. 2). Consequently, it seems that two transient subgroups between the

sea and the gulf herring exists there, while one of them is a subgroup of the sea herring, the other is more similar to the gulf herring.

By the method described in Appendix 2, the location of all points, representing individual fish, was found in the plane coordinate system for all the three sets of characters (see above). The results obtained by this method are in good accordance with the above presented ones. The individuals belonging to the five larger groups distinguished by the method described in the Appendix 1, are situated close to one another also on the graphs composed by the method presented in the Appendix 2. On these graphs separate areas, filled chiefly with the points belonging to groups 1a, 1b, 2, 3a and 3b, can easily be distinguished. On the Figure 3, presented as an example, the position of the points found on the ground of plastic characters, is shown. It can be seen that areas occupied by points belonging to groups 1a, 1b, 2, 3a and 3b, can be well separated.
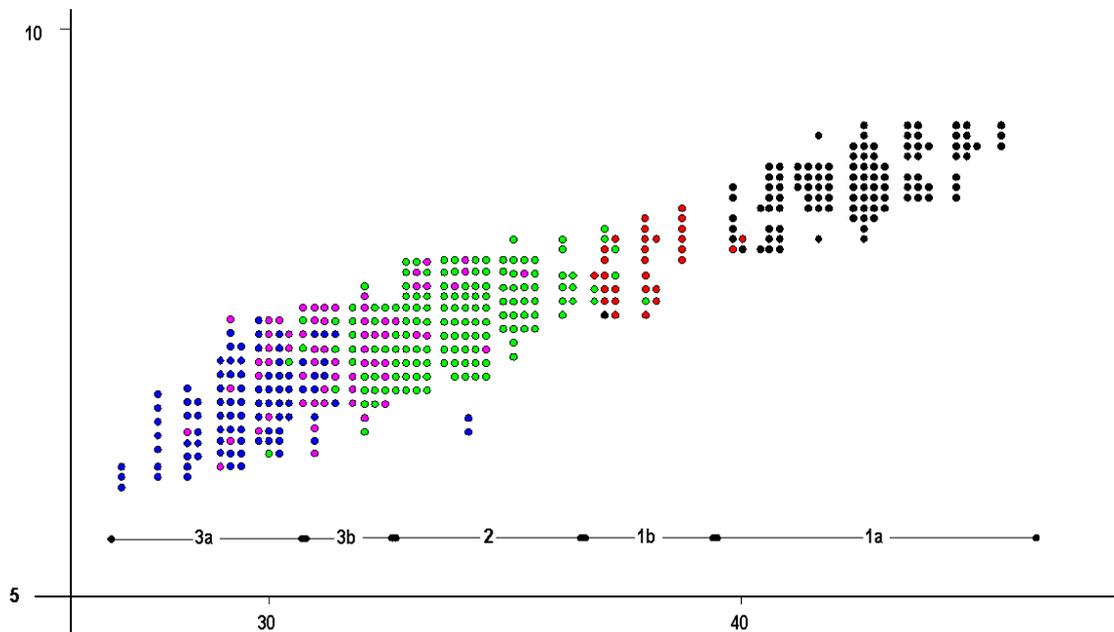


Fig. 3.  Location of points, representing individual fish, determined by the method described in Appendix 2. The members of different groups discerned by the method presented in Appendix 1, are designated as follows: 1a – ● ; 1b – ● ; 2 – ● ; 3a – ● ; 3b – ● .

**Discussion**

In general, it is not possible to discern strictly all the specimen of population of one species. That is caused both by the smoothness of change of environmental conditions in the areas inhabited by these populations and by the reproductive mixing between them. Regarding this, the results obtained show clearly that in the area considered, the autumn herring constitutes a number of groups more or less closely connected with each other. The bigger ones, inhabiting larger and relatively separated areas of the sea, could be treated as populations.

The Swedish south coast herring is located on the graph separately from the others (Fig. 2). No points representing the specimen caught off the Swedish south coast are placed by the computer into the herring groups of the north-eastern Baltic or even into the herring group of the Southern Baltic (subgroup 1b, Table 1). The environmental conditions in the area of distribution of the Swedish south coast herring (the Bornholm Basin) differ considerably from those in the other sampled areas in the vicinity of the Gdansk or Gotland Basin. Therefore we suppose that the Swedish south coast autumn herring constitute a separate population(consisting probably of two parts) differing from the herrings sampled in other locations.

It seems that the herrings caught in the Klaipeda and Liepaja area are more similar to the Swedish south coast herring (they constitute the main bulk of the subgroups 1b but are also present in subgroup 1a) than to the sea herring ring of the north-eastern Baltic. It can be presumed that the Southern Baltic as far as the Liepaja area in the north, is inhabited by the autumn herring hatched mainly on the spawning grounds of the Southern Baltic (owning to inappropriate bottom relief, only a few limited places are suitable for autumn herring spawning off Liepaja). This is favored by the northward current proceeding along the coast in this region. Approximately between the 56° and 57°N the current deviates to the west, finally forming a circular current [7]. Hence it seems that the boundary area between the southern and north-eastern populations of the autumn spawning sea herring is situated in theses latitudes (between Liepaja and Ventspils) and

we are justified to presume that the boundary is connected with the branch of the current deviating westward and limiting the maximum of herring populations of the areas to the north and south of this boundary.

The populations of the autumn herring inhabiting the north-eastern Baltic can be divided into two sets: the populations of the gulf and these of the sea herring. On the graph the subgroup consisting of the gulf herring (3a, Fig. 2) is well separated from other groups. In comparison with other subgroups, this subgroup is denser. Relatively few herrings caught in the Gulf of Riga and the eastern sampling place of the Gulf of Finland, are placed by the computer into the groups other than third. To show the tendency in differences between some important characters of the groups shown on Figure 2 and to compare our data those by other authors, the arithmetic means of these characters are presented in Table 2. It can be seen that in comparison with the sea herring the autumn gulf herring has considerably smaller length, larger eyes (that was stated already by Suvorov [8]) and relatively larger otoliths.

Table 2

Arithmetic means of some morphological characters of the autumn herring of main groups shown in Figure 2; I – $l_s$; II – the ratio $\left(\text{otolith length}/\, l_s\right)$ in per milles; III – the ratio $\left(\text{eye diameter}/\, \text{head length}\right)$ in per cents.

| No. of Sampling Places on Fig. 1 | Groups | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1a | | | 1b | | | 2 | | | 3 | | |
| | I | II | III | I | II | III | I | II | III | I | II | III |
| 1 | 18.4 | 18.8 | 24.5 | | | | | | | | | |
| 2 | 17.8 | 19.3 | 24.3 | 16.5 | 19.3 | 25.0 | 15.2 | 20.3 | 25.3 | | | |
| 3 | 17.6 | 19.0 | 24.7 | 16.2 | 20.1 | 25.1 | | | | | | |
| 4+5 | | | | 16.0 | 20.0 | 24.6 | 14.6 | 20.2 | 25.3 | 13.7 | 20.9 | 25.8 |
| 6 | | | | 16.0 | 19.7 | 25.3 | 14.4 | 20.7 | 26.6 | 13.6 | 20.9 | 26.1 |
| 7 | | | | | | | 14.4 | 20.4 | 25.2 | 12.9 | 21.7 | 26.1 |
| 8 | | | | | | | 13.1 | 21.4 | 25.6 | 12.1 | 21.9 | 26.8 |
| 9+10 | | | | 16.3 | 20.7 | 24.4 | 14.1 | 20.8 | 25.9 | 13.1 | 21.4 | 26.1 |
| 11+12 | | | | | | | 13.4 | 21.7 | 26.6 | 12.8 | 22.1 | 26.8 |
| Average | 18.3 | 18.9 | 24.5 | 16.2 | 19.8 | 24.9 | 14.4 | 20.6 | 25.6 | 12.8 | 21.6 | 26.4 |

The most numerous autumn spawning gulf herring population lives in the Gulf of Riga. There it constitutes a subgroup (3a′, Fig. 2) differing somewhat from other gulf herrings. Also, in the Gulf of Finland and in the vicinity of Hiiumaa, the gulf herring can be found. It is possible that the gulf herring inhabiting the Hiiumaa area, is connected with the population of the Gulf of Riga (the herring larvae hatched in the Gulf of Riga and in the Muhu Sound can be carried by currents through the sounds to the Hiiumaa area). It seems that due to a relatively small abundance of the autumn herring population of the Gulf of Finland (it populates the most north-eastern boundary area of the autumn herring distribution and its numbers are limited by the low survival of larvae in severe wintering conditions) the proportion of the transient subgroup in this gulf is more important than in the Gulf of Riga.

The autumn spawning sea herring of the north-eastern Baltic is a heterogeneous group. The existence of a comparatively numerous of herring (2, Table 1) common for large area from Ventspils to Hiiumaa, hints to a considerable mixing of the autumn herring of this area. Therefore it seems that the sea herring of this area can be treated as a population. But this population includes also subgroups 2‴ and 2⁗ (Fig. 2) characteristic of the Irben-Soela and Ventspils area respectively (Table 1). It means that in various parts of the autumn spawning se herring can be found. These could be considered as the nuclei of subpopulations. The existence of such a nucleus of a subpopulation in the Ventspils area can be connected with the spawning places in the Irben Sound, in the Gulf of Riga and with those to the west of Saaremaa.

## Summary

Considering the results of the above investigation, the degree of geographical and oceanological isolation of various parts of the Baltic Sea as well as the level and fluctuation of the abundance of autumn herring different grounds, the following autumn herring populations can be distinguished in the areas examined:

1. The Swedish south coast sea herring consisting probably of two subunits.

2. The southern Baltic sea herring with its northern boundary reaching the area between 56° and 57°N. We have no samples covering satisfactorily the areas inhabited by the above populations, therefore it is possible that our conclusions are valid for a part of these groups only.

3. The sea herring of the north-eastern Baltic occurring from the area between the 56° and 57°N to the mouth of the Gulf of Finland. This population consists of at least of two subpopulations connected probably with different spawning grounds.

4. The Gulf of Riga herring.

5. The Hiiumaa gulf herring population, possibly connected with the population of the Gulf of Riga herring.

6. The small autumn herring population in the middle and eastern part of the Gulf of Finland.

## References

1. Hessle, Shr., The herrings along the Baltic coast of Sweden. – "Publ. De Circonst," 1925, No. 89, pp. 1-57.

2. Popiel, J., Differentiation of the biological groups of herring in the Baltic. – "Rapp. Et Proc.-Verb.," 1958, vol. 143, II, pp. 114-121.

3. Rannak, L., Kevadudüräime biologiilisi rühtumusi otoliitide põhjal." – Eesti NSV TA Toimet. Biol. Seeria," 1967, 16, No. 1, lk.. 41-53.

4. Rauck, G., The structure of otoliths from the Baltic herrings – a helpful means for the separation of biological groups. – "ICES CM Herring Comm. Paper,", 1-65, No. 39, pp. 1-4.

5. Strzyzewska, K., Stadium porownawcze populacji sledzi tracych sie u Polskich webreźy Baltyku. – "Prace MIR,", 1969, tom. 15, Seria A, pp. 211-277.

6. Оявеер Э. А., О Различении сезонных рас салаки северо-восточной части Балтийского моря по отолитам. – «Изв. АН ЭССР. Сер. Биол.,» 1962, т. II, вып. 3, с. 193-208.

7. Соскин И. И., Кузнецова Л. Н., Соловьев В. И., Течения Балтийского моря на основе обработки гидрологических наблюдений динамическим методом. – «Тр. ГОИН'а,» 1963, вып. 73, с. 76-95.

8. Суворов Е. К., К ихтиофауне Балтийского моря. – «Тр. Балт. Эксп.,» 1913, вып. 2, с. 37-99.

## Appendix I

There are many intuitive ideas, often contradicting, of what is a cluster. Although it is rather difficult to develop exact mathematical formulation of the cluster separation task there are authors who think that clustering techniques are already established and the need is simply for more correctly analyzed data. The real examples are as a rule quite badly structured and often the formal techniques fail on such data where the classification is already known. This situation demonstrates the basic failure of many techniques in numerical taxonomy.

As usually we describe every object by a vector of measurements $\langle x_1, x_2, ..., x_k \rangle$. For every pair of objects $E_i$ and $E_j$ we define a distance $d_{ij}$ between those objects

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + ... + (x_{ik} - x_{jk})^2} \tag{1}$$

(usually all measurements are standardized beforehand).

For $N$ objects we can calculate a full matrix of distances

$$D = \begin{Vmatrix} 0 & d_{12} & d_{13} & . & . & d_{1k} \\ d_{21} & 0 & d_{23} & . & . & d_{2k} \\ . & . & . & . & . & . & . & . & . \\ d_{k1} & d_{k2} & . & . & . & d_{kk} \end{Vmatrix} \tag{2}$$

Many empirical studies have used the matrix of distances to find clusters on the set $\{E_1, ..., E_i, ..., E_k\}$.

In this section we describe a new effective method for clustering, which uses some ideas of the theory of graphs. As the first step we emphasize that the whole matrix of distances is rarely needed to disclose the structure of the system of objects. Therefore, for every object, we take into account not more than $M$ of nearest neighbors.

As an example let us have a system of 9 objects (Fig. 4) with their interconnections – edges. The matrix of nearest neighbors for our graph is as follows:

$$MND = \begin{Vmatrix} 5(1) & 6(1) & 3(2) & 0 & 0 & 0 \\ 4(1) & 3(2) & 7(3) & 0 & 0 & 0 \\ 4(1) & 5(1) & 1(2) & 2(2) & 0 & 0 \\ 2(1) & 3(1) & 5(1) & 7(3) & 0 & 0 \\ 1(1) & 3(1) & 4(1) & 6(1) & 7(3) & 0 \\ 1(1) & 5(1) & 7(3) & 0 & 0 & 0 \\ 2(3) & 4(3) & 5(3) & 6(3) & 8(3) & 9(3) \\ 7(3) & 9(3) & 0 & 0 & 0 & 0 \\ 7(3) & 8(3) & 0 & 0 & 0 & 0 \end{Vmatrix}$$
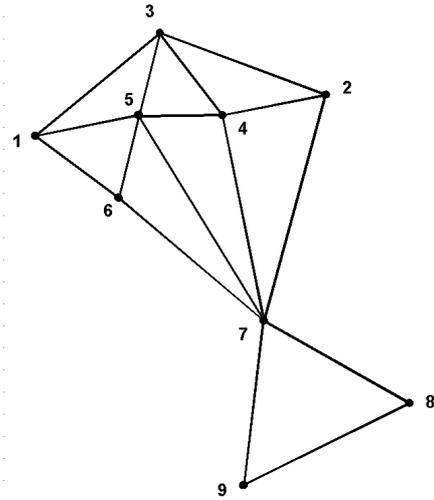


Fig. 4

It is easy to check that every row $i$ of that matrix contains a list of objects $j$ directly connected with a given object $E_i$. In brackets the distances $d_{ij}$ are given. We will denote the matrix of nearest neighbor distances by *MND*.

Usually we need data about 8-10 nearest neighbors only. The economy in memory space is very important, e.g., in the case of 1,000 objects we need only 10,000 memory locations and not 500,000 as in the case of the full matrix.

We will use *MND* as a starting point to create some not very complicated mathematical constructions.

Let $W$ be the list of edges (pairs of objects) in the *MND*. For every edge $e = [a,b]$ we will define a subset $W_b^a$ of the list $W$ as follows.

**Definition 1**. Subset $W_b^a$ of $W$ is called a proximity space of edge $[a,b]$ if

    a)     for every pair of objects $x$ and $y$, which are connected at least with one edge in $W_b^a$, there exists a path joining $x$ and $y$, and

    b)     every edge, which is a member of that path belongs to the subset $W_b^a$.

In terms of graph theory proximity space is a subgraph connected with the edge $[a,b]$.

Example. Let us take the edge $[4,5]$ on Fig. 4. Its proximity space is $W_5^4$ is the subgraph $W_5^4 = \{[3,4],[3,5],[4,7],[5,7],[2,4],[1,5],[5,6],[4,5]\}$.

**Definition 2**. The system of proximity spaces is called the proximity structure if for each edge $w = [a,b]$ there exists nonempty $W_b^a$ in the system.

Sometimes it is useful to exclude the edge $[a,b]$ from the proximity space $W_b^a$. We will denote this as follows: $W_b^a \setminus [a,b]$ and call the result a reduced proximity space.

So far, for every edge $[a,b]$, we took into account only the value of the distance $d[a,b]$ between $[a,b]$. In what follows it is useful to introduce a new notation. We can assign a real number (weight $\pi$), which is different from the distance to every edge on the graph. For example, let us define the weight of the edge for every edge on Fig. 4 as

$$\pi[x,y] = d[x,y] + r[x,y],$$

where $d[x,y]$ is the Euclidean distance (1) between $x,y$, and $r[x,y]$ is the number of triangles, which can be built on the edge $[x,y]$. For example $\pi[4,7] = 3 + 2$, $\pi[7,8] = 3 + 1$.

Let us suppose that for a graph $W$ there is given its proximity structure $\mathcal{L}$ and let $f(x)$ be a real function.

**Definition 3**. The function $f_b^a(\pi)$ given on all weights of the edges in $W_b^a$ is called the influence function of the proximity structure $\mathcal{L}$ if it is provided that

$$f_a^b(\pi[x,y]) \leq \pi[x,y]$$

for each $[x,y] \in W_b^a \setminus [a,b]$, where $\pi[x,y]$ is the weight of the edge $[x,y]$.

In other words for every edge $[x,y]$ we can find a new weight in the reduced proximity space $W_b^a \setminus [a,b]$

$$\pi'[x,y] = f_b^a(\pi[x,y]). \tag{3}$$

To understand the idea of the influence function better let us again use the graph on Fig. 4. The influence function represents the value of the number of triangles after the elimination of the edge $[a,b] \in W_b^a$ from the list $W_b^a$, e.g., let us take again the set $W_5^4$, then

$$f_5^4(\pi[3,4]) = f_5^4((d_{34} + r_{34}) = (1+1)) = (d_{34} + r_{34}') = (1+0) = 1\,;$$
$$f_5^4(\pi[3,4]) = f_5^4((d_{56} + r_{56}) = (1+0)) = (d_{34} + r_{34}') = (1+0) = 1\,;$$
$$f_5^4(\pi[3,4]) = f_5^4((d_{47} + r_{47}) = (3+1)) = (d_{34} + r_{34}') = (3+0) = 3\,.$$

$$MNW = \begin{Vmatrix} 5(3) & 6(2) & 3(3) & 0 & 0 & 0 \\ 4(3) & 3(3) & 7(4) & 0 & 0 & 0 \\ 4(3) & 5(3) & 1(3) & 2(3) & 0 & 0 \\ 2(3) & 3(3) & 5(3) & 7(5) & 0 & 0 \\ 1(3) & 3(3) & 4(3) & 6(3) & 7(5) & 0 \\ 1(2) & 5(3) & 7(4) & 0 & 0 & 0 \\ 2(4) & 4(5) & 5(5) & 6(4) & 8(4) & 9(4) \\ 7(4) & 9(4) & 0 & 0 & 0 & 0 \\ 7(4) & 8(4) & 0 & 0 & 0 & 0 \end{Vmatrix}$$

After having understood the influence function of an edge we can easily find the set of new weights for a whole subset $H \in W$. Let us look at the set $\overline{H} = W \setminus H$ and arrange its edges in some order $\langle e_1, e_2, ... \rangle$. We shall find the proximity spaces of the edges in $\langle e_1, e_2, ... \rangle$ and use the formula (3) recursively.

Now we are ready to introduce our algorithm to discover the data structure.

Suppose that the selection of the proximity structure and the influence function has been done. Use the following algorithm.

A1. Find the edge with the minimum weight and store its value.

A2. Eliminate the edge from the list of all edges and compute the weights for proximity spaces of the minimal edge using the recursive procedure (3).

A3. Go through the list of edges and find the first edge with the weight less or equal to the stored weight. Go to A2 to eliminate that edge. If there is no such edge go to A4.

A4. Check whether there are any more edges in $W$. If yes go to A1, if not stop the calculations.

We will demonstrate how the algorithm works on our graph on Fig. 4.

We will define weights for all edges by

$$\pi[x, y] = d[x, y] + r[x, y].$$

First of all, we compute the matrix of weights using the matrix of distances (2)

We will follow the algorithm through all stages.

A1. Minimal edge is $[1,6]$ with weight $\pi[1,6] = 2$. To store its value let $u = 2$.

A2. We eliminate edge $[1,6]$ from the list $W$ and therefore we have to change the weights of $W_6^1 \setminus [1,6]$: $\pi'[1,3] = 3$; $\pi'[1,5] = 2$; $\pi'[5,6] = 2$; $\pi'[6,7] = 4$.

A3. The first edge with the weight less or equal to $u$ is the edge $[1,5]$. Now we return to step A2. After 9 steps with $u = 2$ we have got the sequence of edges: $\langle [1,6], [1,5], [1,3], [3,5], [3,4], [2,4], [2,3], [4,5], [5,6] \rangle$.

Now we to take $u = 3$, and we get $\langle [2,7], [4,7], [5,7], [6,7] \rangle$. At last $u = 4$ and we get $\langle [7,8], [7,9], [8,9] \rangle$.

It is easy to check that those ordered lists of edges represent the structure of our graph very nicely.

For graphical output we construct a connected tree from the ordered edges (a tree is a graph without circles).

We will follow the ordered lists of edges leaving out all those edges $[a,b]$ both ends $a$ and $b$ of which are already in the list.

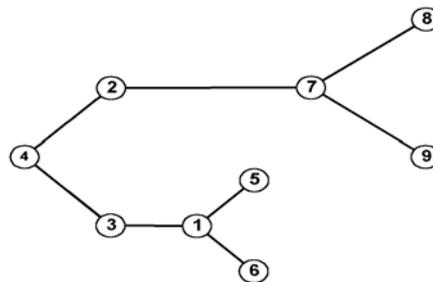We get a sequence $\langle [1,6], [1,5], [1,3], [3,4], [2,4], [2,7], [7,8], [7,9] \rangle$ and construct the tree



Fig. 5

How do we know that object number 4 is on the top of the tree? We simply denote the number $S(x,y)$ of steps needed to reach the point $y$ from the point $x$ on the tree (f.e. $S(1,2) = 3$, $S(1,8) = 5$) and for every object $x$ we then locate the object, which need the maximum number of steps. Afterwards, as the top of the tree, we will take the object for which that maximum is minimum.

With the reference to real data a tree created following just explained rule is presented on Fig. 2.

**Appendix 2**.

To analyze effectively large dimensional data matrices with hundreds of objects and tens of attributes we represent a new method, which has recommended itself in practice.

Let every object $e_i$ have attributes $x_1, x_{2,}..., x_k$. All data about the system of objects $\{e_i\}$ can be collected into data matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & .... & x_{1k} \\ x_{21} & x_{22} & .... & x_{2k} \\ ... & ... & .... & ... \\ x_{N1} & x_{N2} & .... & x_{Nk} \end{pmatrix}$$

Let us study linear functions of attributes

$$L = \sum_{i=1}^{k} c_i \cdot y_i \quad \left( \sum_{i=1}^{k} c_i^2 \neq 0 \right).$$

One of the most well-known linear functions of attributes is the arithmetical mean

$$L = \frac{1}{k} \cdot \sum_{i=1}^{k} y_i .$$

Two linear functions $L = \sum_{i=1}^{k} \ell_i \cdot y_i$ and $M = \sum_{i=1}^{k} m_i \cdot y_i$ are independent (orthogonal) if $\sum_{i=1}^{k} \ell_i \cdot m_i = 0$ . A linear function is called a contrast if it is orthogonal to the arithmetical mean. For simplicity let us suppose that we have only four attributes for one object. Then we have the coefficients of the mean and the 3 possible contrasts gathered into a matrix as follows:

$$H = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{pmatrix}.$$

Every other contrast of the observations can be expressed as a linear function of those four vectors. For every vector $x = \langle x_1, x_2, ..., x_4 \rangle$ we can find the values of its 4 independent linear functions by simply using the definition of contrasts. For example the vector $\langle 3, 2, 4, 2 \rangle$ corresponds to the vector $\langle 11, 3, -1, -1 \rangle$.

Such contrasts do have a very important property.

If the number of attributes is a power of 2, the calculations needed to find the contrasts are simplified considerably. We represent here the scheme for $k = 4 = 2^2$.

H1. We find the sums and differences of attributes by pairs.

$$x_1 \quad z_1' = x_1 + x_2$$
$$x_2 \quad z_2' = x_3 + x_4$$
$$x_3 \quad z_3' = x_1 - x_2$$
$$x_4 \quad z_4' = x_1 + x_2.$$

H2. We repeat the process $log_2(k-1)$ times (therefore in the case of 4 attributes once more).

$$z_1' \quad z_1' + z_2' = x_1 + x_2 + x_3 + x_4 = z_1''$$
$$z_2' \quad z_3' + z_4' = x_1 - x_2 + x_3 - x_4 = z_2''$$
$$z_3' \quad z_1' - z_2' = x_1 + x_2 - x_3 - x_4 = z_3''$$
$$z_4' \quad z_3' + z_4' = x_1 - x_2 - x_3 + x_4 = z_4''.$$

The $z_i$-s are the values of contrasts. For example let us take the same vector as before

$$
\begin{array}{ccc}
3 & 5 & 11 \\
2 & 6 & 3 \\
4 & 1 & -1 \\
2 & 2 & -1
\end{array}
$$

Performing the process for every object we get in place of our raw data matrix $X$ the matrix of contrasts $Z$.

The process of calculating the values of $Z$ is quite easy to perform even by hand. In case the number of attributes is not a power of 2 we can take the missing attributes equal to zero.

The next step is to choose such $z_i$-s from the contrasts, which are most important in the sense that they are changing most violently in our system of objects. Those contrasts will show us the most important relation between the objects.

As a criterion we will use the sum of squares of the contrasts

$$S_\ell = \sum_{j=1}^{N} z_{\ell j}^{\ell} \;.$$

If we choose two most important constraints, we can easily draw up the map using those contrasts as coordinates of objects.

The process of choosing two best contrasts provides us with the possibility of estimating the goodness of approximation. We have to find only the quotient of the sum of two maximal $S_i - S$ to the sum of all squares

$$Q = \frac{S_{i\,\max} + S_{i\,\mathrm{submax}}}{\displaystyle\sum_{i=1}^{k} S_i} \;.$$

The quotient $Q$ tells us how much of the whole variation is described by those two contrasts.

In case we need three axes instead of two, it is possible to draw a twodimensional picture. Let the needed three contrasts be $z_{\ell 1}$, $z_{\ell 2}$ and $z_{\ell 3}$. We have to use only the new coordinates

$$X_1 = arctan \frac{z_{\ell 2}}{z_{\ell 1}} \;;$$

$$X_2 = arctan \frac{z_{\ell 3}}{\sqrt{z_{\ell 1}^2 + z_{\ell 2}^2}} \;.$$

The organization of computations on the computer is quite effective. We can store all objects on the magnetic tape or drum, bring them sequentially into core memory, compute the contrasts, simultaneously also compute the square of contrasts and gather the sums of squares $S_i$ in $k$ storage locations. All calculated contrasts would be stored on magnetic tape.

We have to order the sums $S_i$, to analyze how good an approximation we can get with two and three contrasts and to plot the objects on the line printer.

The results of the method are presented on Fig. 3.